DOCUMENT RESUME

ED 133 364 TM 005 979

AUTHOR Moonan, William J.
TITLE PRIME: Personnel Research Instrumentalities for Measurement and Evaluation.
PUB DATE [Sep 75]
NOTE 21p.; Paper presented at the Annual Meeting of the Military Testing Association (17th, Fort Benjamin Harrison, Indiana, September 15-19, 1975); For a related document, see EL 128 402

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Computer Programs; *Mathematical Models; Measurement Techniques; *Personnel Evaluation

ABSTRACT

The purposes of the paper were to develop a structure by which assessment models could be uniquely characterized and to develop assessment models and their associated computer programs which could be used for personnel research purposes. To these ends two aspects of assessment, measurement and evaluation, were identified. Their distinction was mainly in the process used to respond to the items of an assessment instrument. The parameters of the assessment models were identified. These were variously combined and operationalized into either measurement or evaluation models. The intention of this process was to make evaluation more measurement-like by using and developing evaluation models, for polychotomous response information, which were patterned after Rasch latent-trait measurement models. Also developed were some special models for use in decision making paradigms and for assessment situations for which an estimation sample does or does not exist. Analysis of these models and their associated data were provided by a system of computer programs called PRIME. In all, five programs were developed, illustrated, and documented. Applications of these models and computer programs to personnel research areas of the Navy Personnel Research and Development Center research program was discussed. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED133564

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PRIME: Personnel Research Instrumentalities for Measurement and Evaluation

William J. Moonan, Ph.D.
Navy Personnel Research and Development Center

Introduction

The purpose of this report is to develop the framework for a structure of "assessment" and to provide procedures for the analysis of "assessment" models developed from this structure. The author feels that this endeavor is necessary because of the apparent confusion that the topic of assessment appears to have on the minds of many involved in personnel research activities. We shall proceed by examining the fundamental aspects of assessment procedures, developing assessment models, and developing computer programs for the analysis of these models.

Background

The Topic of this report is assessment for personnel R&D projects. The author assumes that the reader is in agreement with him that the framework of assessment of personnel R&D in a frightful mess and needs clarification, c.f. Anderson's (1973) Anatomy of Evaluation, which contains a very large list of articles, concepts, definitions and techniques relating to one aspect of assessment, the Page (1975). It is assumed that the author and reader are in agreement that the topic needs simplification and clarification with respect to the specification and definition of its fundamental aspects, and that the practice of personnel assessment can be improved by postulating in simple structural framework and by the development and use of formal assessment models.

The author further assumed that the reader has had a professional education with some training on measurement and statistical topics and is motivated toward improved personnel R&D and wishes to learn of a new attempt to characterize a simple and meaningful structure for the topic under discussion, and, at the same time, to learn of some new assessment models and their associated computer programs which can be used on personnel R&D work.

Types of Assessment

A. Aspects of Assessment

By the term "personnel assessment" the author refers to the "measurement" or "evaluation" of personnel. The distinction between these pro-

TM005 979

cesses will ultimately be realized by a technical description. We are assuming also that the "object" being assessed is usually a person, since we are dealing primarily with persons in personnel research work. Our procedures are, upon occasion, also applicable for assessing other types of objects or entities such as systems, organizations, operations, programs, tasks, etc., but we will not emphasize these applications. Our purpose primarily being to assess personnel.

The substructures, assessment and evaluation, each has an enormous literature, not to be reviewed here. The subtopic "measurement" has undergone a long and steady evolution, culminating in various latent trait models and conjoint measurement procedures. The subtopic "evaluation" is more diffuse and less mathematically structured, and has a literature in the education and psychology fields primarily related to evaluations of complex programs rather than personnel. These evaluations frequently consisting of measurements or their functions. We shall not pursue this course in our development, rather we shall tend to make the process of personnel evaluation have analysis procedures similar to those of measurement.

B. Elements of Assessment

We have already identified the object or entity of assessment as one of the elements of this topic. Another element is the assessment instrument consisting of assessment items, each of which is associated with a set of item response categories. The assessment instrument in measurement situations is a test or examination whereas the instrument in evaluation situations is usually either a questionnaire, checklist or inventory. The response categories of the items of a test may be "multiple choice." For our purposes we consider that each item has only two operational responses categories, i.e., correct or incorrect. Evaluation instruments usually also have multiple category responses, which we assume are also expressed as levels of quality, several rather than two.

We postulate the existence of another assessment entity to which we append the label "assessor," whose function it is to assess the object of assessment, who is or isn't another person. In measurement the person is using his own latent abilities or competencies to designate the response categories of the items of the assessment instrument. Whereas for evaluation situations, the assessor usually utilizes his subjective judgment about himself or another entity to designate the appropriate response categories of the items of the assessment instrument.

As promised the previous two paragraphs provide our distinction between the concepts of measurement and evaluation. We summarize the various elements of assessment as follows in Exhibit 1. We realize, in truth, that our distinction between "measurement" and "evaluation" is not as sharp as we would like to make it. More realistically we may consider

EXHIBIT 1

Elements of the Assessment Process

	Assessment Type	
Assessment Entity	Measurement	Evaluation
Assessee	testee	person/entity
Assessor	testee	evaluator
Item response designator	test, examination	evaluator
Assessment instrument	problem solving,	questionnaire,
Assessment items	recognition or recall items measuring latent abilities, or items of interest and attitude	checklist statements or questions of status, opinion, value, skills and achievement
Item response categories	dichotomous; correct or incorrect	polychotomous; levels of quality, desirability or value
Essential Process	use of latent abilities	subjective judgment

"Assessment" to be a bipolar concept with some assessment situations more measurement-like than evaluation-like or vice versa. Fortunately the distinction is not critical and an appropriate assessment model for some R&D assessment project can easily be located among those we shall propose.

In addition to these factors, we consider some further elements of importance. One of these is called the Assessment Sample, which consists of those entities to be assessed once the assessment procedure has been formalized after an analysis. This analysis is usually made on data provided by an Estimation Sample from which the parameters of the assessment model are estimated and the scoring procedure formalized. The purpose of the analysis is to develop the scoring procedure to be used for providing quantified magnitudes estimating abilities or competencies in measurement situations or the "values" of the assessees in evaluation situations. Other purposes are to analyze the assessment instrument items and to examine the "fit" of the response data to the assessment model.

C. Assessment Model Spectrum

In order to characterize the assessment system proposed herein, we wish to define a set of symbols and their definitions which collectively we designate as the Assessment Model Spectrum. The purpose of this spectrum is to provide a simple structural characterization of an assessment problem and provide for the use of this spectrum for identifying the appropriate assessment model analysis. Our spectrum consists of the following nine parameters each having one or more corresponding designators as shown in Exhibit 2.

The first parameter, Aspect, indicates that the assessor is considering the assessment problem to be either a measurement problem or an evaluation problem in the sense described previously. Our assessment procedure assumes that the assessor is assessing basic characteristics or traits that are not directly observable and therefore these traits are referred to as latent traits, or he is evaluating some value. It is the purpose of the process to estimate a person's standing on a value or trait, which can in turn be used for other operational or research purposes. In measurement situations, these traits are referred to as abilities, whereas for evaluation situations, they are referred to as values, in each case being further characterized by the name of the ability or value being assessed. This name being listed as the second assessment spectrum parameter.

The parameter "instrument" refers to the type of assessment instrument utilized, most frequent types being tests, questionnaires, or checklists. The number of items on the assessment instrument is listed in the fifth parameter. Our procedures require the assessment instrument, no matter

EXHIBIT 2

Assessment Model Spectrum Parameters and Subparameters

Nr.	Parameter Name	Subparameter
1.	Aspect	M = measurement (primarily) E = evaluation (primarily)
2.	Trait	L = latent trait name to be measured V = value evaluated
3.	Entity	P = person S = system O = object or other
4.	Instrument	T = test or examination Q = questionnaire C = checklist O = other
5.	Items	I = number of assessment instrument items
6.	Responses	M = number of response categories
7.	Application Sample	N = size of application sample
8.	Estimation Sample	E = size of estimation sample
9.	Assessment Process	A = latent ability or competency S = subjective judgment

what its form, to consist of a subset of statements, questions, or whatever, which may be labeled as items. Each item is associated with a set of common response categories, the number of which we describe by the parameter M. If M = 2, the response categories are described as binary or dichotomous, where if M > 2, the response category type is called polychotomous. In multiple choice tests, M = 2, since the response is judged correct or incorrect, in spite of the fact that the testee has more than two alternatives from which to choose for each item. Our concern is for the number of "operational" response categories to be used for analysis purposes and not for the number of instrumental response categories. In our assessment model system, the response categories for evaluation instruments represent levels of quality possibly realized by the assessee relative to the items' substantive content. For example, for the item "How would you rate this instructor with respect to his general teaching ability?", for which we may have the response categories:

poor
fair
adequate
good, or
excellent

Finally, we are concerned with the existence and size of two sets of data resulting from persons for whom response information may be available. The first sample, called the Application sample is not ordinarily used for analysis. It is this sample, the members of which are to be assessed, once the assessment procedure is defined. In some cases, this application sample is not immediately available, since the assessment procedure is being developed and to be applied some time in the future. The other sample, called the Estimation sample, if it exists, provides the response data to be utilized for it estimating the parameters of the assessment model. Part of our assessment system provides models to be utilized for those occasions, usually occurring in evaluation situations, for which an Estimation sample does or cannot exist because of operational constraints. There are occasions, usually in measurement situations, when the Application and Estimation samples are identical. This circumstance causes no difficulty for the models of our system. For most measurement model analysis, we require an Estimation sample of adequate size.

The parameter "entity" is either a person, system, object or some entity which can be observed producing or made to produce information which can be transformed into assessment instrument response categories.

D. Models of Assessment

The modern scientific approach to measurement theory may be said to have begun with the publication of Lord's (1952) "A Theory of Test Scores"

wherein the fundamental concept of "true scores" was postulated. A true score model may be characterized by the equation

(1) $X = T + E$ where

X represents the observed test score, T represents the true score, and E represents random error. Lord extensively published research work on this and related true score models during the score of years 1950-1970.

A related model, which for lack of a better name, we shall call the "Minnesota" model,* is

(2) $R = A + I + e$ where

R represents the response made to one of the items of a test. A represents the ability of the item respondent, I represents the item effect, and e is the random error associated with the item response. In our opinion, (2) represents the measurement model commonly assumed to exist in the minds of many personnel research workers, although its analysis is not carried forth in the popular measurement textbooks, including Gulliksen's (1950), in the fields of education and psychology. Practitioners intending to utilize the concept of (2) usually proceed to construct a measuring device whose item responses are then judged as correct or incorrect and which are accumulated into total scores. These scores are not for the purpose of estimating the ability of the testee as would be the objective if the practitioner were interested in measurement, but the scores are frequently utilized for relating them to other information and as such represent statistical, rather than measurement, objectives. As a consequence, we attach the terms "pseudo-measurement" or quasi-measurement" to this practice.

Various classes of latent-trait models are described by Lord and Novik (1968). Distinctions of these models are accomplished by denoting the parameters involved in the model specification. These parameters may be listed in Exhibit 3.

Some or all of these parameters are assembled into a mathematical expression describing the probability of a given type of r(n,i) as a function of the parameters used in the model. This model is then said to be a latent trait measurement model. The so-called Birnbaum (1968) three-item parameter model utilizes the three-item parameters δ (i), d(i), and g(i) as well as α (n) in developing an expression for the probability of a correct response. The Birnbaum (1968) two-item parameter model involves the two-item parameters of δ (i) and d(i), as well as α (n), for the same purpose.

---

*The label "Minnesota" is appropriate since Professors Palmer O. Johnson, Cyril J. Hoyt, and others used this model, during 1940-1960, as an analysis of variance model to instruct students and carry out, and develop various psychometric analyses at the University of Minnesota.

EXHIBIT 3

Parameters of Assessment Models

α (n):	a real number representing the ability or latent trait of the n^{th} person.
δ (i):	a real number representing the "difficulty" of the i^{th} item.
d (i):	a real number representing the i^{th} item's discrimination power.
r (n, i):	the response made to the i^{th} item by the n^{th} person.
g (i):	a real number representing the effect of guessing the correct answer to the i^{th} item.

Rasch (1960) also developed a variety of latent trait models one of which, the Binary Measurement Model, BMM, develops the expression for the probability of a correct response as a function of $\delta(i)$ and $\alpha(n)$ only. He was able to show that certain desired measuring properties called "objectivity" resulted from this specification. Briefly, objectivity, means that we can achieve both

(a) person-free item measurement, and
(b) item-free person measurement by this specification.

The Rasch model does not actually utilize $\delta(i)$, but rather, another parameter designated as $\varepsilon(i)$ which we may assume is conceptually and inversely related to $\delta(i)$ and referred to as "item easiness."

In our assessment system, we shall utilize the BMM as the basis for measurement and shall utilize generalizations of this model for polychotomous responses as the basis for making quantitative evaluations. We do this to maintain a degree of continuity of process, and for the reason that the Rasch models represent the simplist latent trait model and therefore can be analyzed mathematically and computationally with minimal skill and effort.

We conclude this section by indicating the assumptions required by the BMM. These are:

(1) Test items are scored dichotomously.

(2) The probability of a correct response is only a function of the ability, $\alpha(n)$, of the respondent and the easiness, $\varepsilon(i)$, of the item.

(3) Responses to the items are stochastically independent given $\alpha(n)$ and $\varepsilon(i)$.

The Prime System

A. Introduction to PRIME

Any assessment is the result of measuring or judging something. That is, applying some assessment process or procedure, some number or code is recorded as a result. The purpose of making an assessment is to use it for doing something. Hence the object of making assessments is to provide a basis for action.

If the assessment is related to mental measurement, the "assessment process" is usually an intellectual function, such as problem solving or the use of some memory process. In these kinds of situations, the "assessment process" is carried out by the person (entity) being measured.

If the assessment is related to evaluation, the "assessment operation" (a judgment) is often carried out by an evaluator with respect to the entity being evaluated.

In either event, the number or code resulting from assessment is a specification of a response category, i.e., a multiple choice selection or designation of a value category of assessment instrument items.

As the acronym "PRIME" indicates, it is a system devoted to "instrumentalities" for making measurements and evaluations in personnel research problems. These instrumentalities represent measurement/evaluation models and their associated computer programs designed to operate on computers available to NPRDC personnel:

The PRIME programs are enumerated as follows:

FRM:	The Frequency Ratio Method (Binary Measurement Model)
TIM:	The Inspection Model
PAM:	The Polychotomous Assessment Model
PMM:	The Polychotomous Measurement Model
HELP:	The Handy, Empirical Little Prioritizer
(MAXKO):	Auxiliary PRIME Program

The objective of the PRIME system is to bring "order" into the topic of assessment and to provide models, and their associated computer programs, to workers in the field of Personnel Research.

We next provide a brief overview of each instrumentality and ultimately indicate an application for it in Personnel Research Programs.

B. Briefs of PRIME Instrumentalities

FRM: The Frequency Ratio Method. This program is related to the Rasch Binary Measurement Model (BMM) discussed by Moonan (1969, 1974) and represents a new and simpler method for estimating the "easiness" of measuring instrument items. The program is offered as a possible substitute for a more extensive IBM 360/65 program called "Max" in the NPRDC Computer Program Library. FRM operates on the NPRDC GA 1830 computer system. Documentation and an example using FRM for analyzing BMM data is provided in Moonan and Covher (1975a).

FRM is useful in those measurement situations where the BMM applies and some cognitive domain characteristics, such as intelligence, ability, aptitude or competency is required to be measured by items whose responses are judged operationally to be dichotomous (e.g., correct or incorrect). BMM and FRM are applicable to most personnel measurement problems where the items scoring parameters are 0, 1.

PMM: The Polychotomous Measurement Model. Situations occur in Personnel Research, however, when it is not convenient or desirable to measure an entity property by using dichotomously scored items. Such occasions arise when measuring attitudes, interests, and motivations.

For these purposes we can use a generalization of the Rasch BMM to the polychotomous item response case Rasch (1970). We refer to this model as PMM. Anderson (1973) developed a conditional maximum likelihood computer program for this model. The author has attempted to secure this program for PRIME in the following ways:

(a) card punching a listing of the program as contained in Andersen, et al. (1973);

(b) requesting a program copy from ETS's Data Processing Division;

(c) request a program copy from Allerup, P. and Sorber, G. (1975).

Procedure (a), because of poor printing, produced a program deck that would not compile on NPRDC computers. It turned out, that in spite of Andersen et al. (1973), ETS did not have an operating program either. They offered information about a similar proprietary University of Chicago program. We have requested a PMM copy from Denmark, and have confidence in receiving it but have, as yet, not acquired it because of the delay due to distance.* Nevertheless we include PMM in the PRIME System. The chief advantage of Andersen's PMM program is that it obtains maximum likelihood estimates of the PMM parameters including the scoring parameters of the polychotomous responses to the assessment instrument items.

PMM is useful for measuring interests and attitudes as well as for analyzing evaluative instrument data.

PAM: The Polychotomous Assessment Model. Because of early concern over acquiring Andersen's PMM program, it was decided to develop an approximation program, called PAM, to the PMM program. This model handled polychotomous response data, but reated response categories independently. Rather than estimate the item scoring parameters, it defined them mathematically. The item parameter estimation procedure used was that derived for use with program FRM. Using the scoring parameters, entity scores can be derived and estimates of entity attitude, interest or whatever can be obtained.

Documentation for PAM is contained in Moonan and Covher (1975b) and provides an example using SVIB data collected from naval personnel.

TIM: The Inspection Model. Occasionally in Personnel Research, there occurs a situation wherein it is either infeasible or impossible to collect response information on the items of an assessment instrument

---

*Received 31 July 1975.

from an Estimation sample, yet assessments must be made on an Assessment sample. If the assessments are evaluations and the items are polychotomous, the TIM model is appropriate. The item parameter information is obtained from item "importances" supplied by the evaluator.

Entity "scores" are obtained from a weighted average of the frequencies of item response categories (obtained during evaluation) and item response scoring parameters. A TIM model entity parameter, y, associated with each possible score is provided by the program. Such an estimate provides knowledge of the amount of the quality, measured by the assessment instrument, possessed by the evaluated entity. Please note that TIM is the first PRIME instrumentality discussed which does not require an Estimation sample which provides response data from which item parameters can be estimated. Documentation for TIM is contained in Moonan, et al. (1975c).

HELP: Handy Empirical, Little Prioritizer. In those assessment situations frequently associated with decision making, and others, one is obliged to provide quantified measures of values or utility. Often this process involves the consideration of several factors of importance. HELP calls these factors "Dimensions of Value" and requires the user to provide a consistent ratio scale of the importance (to him) of each Dimension. Each entity of the Assessment sample is then judged as to quality on each Dimension. This information is then synthesized by the HELP model into a quantified measure of entity value. These values then being used with other decision making, statistics or Operations Research models for the purposes at hand.

Such procedures, which quantify multiple-dimensional judgments, are rarely available and few are as simple as HELP. Consequently it is very important for Personnel Research purposes. The HELP documentation is given in Moonan and Covher (1975d).

MAXKO. The purpose of this program is to provide maximum likelihood estimation of the "value" parameter in the TIM and PAM model programs. MAXKO is the polychotomous generalization of subroutine MAXCO used in FRM for the dichotomous response case. The program prints a "scoring table" associating a score derived from a polychotomous assessment instrument with the estimate of the value parameter of the assessment model. The method of estimation is that of iteration on the model equation on p. 2 of Moon and Covher (1975e).

We have now listed and briefly described six PRIME programs.

For clarification purposes the following charts Exhibit 4, and Exhibit 5, characterize the relationship of 4 of the PRIME instrumentalities in Exhibit 4, and the assessment parameters of each model are shown in Exhibit 5.

EXHIBIT 4

Relations of Four PRIME Instrumentalities

Operational Response	Item Response Data	
	Objective	Subjective
Binary	FRM (BMM)	TIM
Polychotomous	PMM	TIM

EXHIBIT 5

Specification of Assessment Parameters for Five PRIME Instrumentalities

Parameter	FRM	TIM	PAM	PMM	HELP
Aspect	M	E	E,M	M	E
Trait	L	V	L,V	L	V
Entity	P	P,S,O	P	P	P,S,O
Instrument	T	Q,C	Q,C	Q,C	Q,C
Responses	M=2	M≥2	M≥2	M≥2	M≥2
Estimation Sample	Yes	No	Yes	Yes	No
Application Sample	Yes	Yes	Yes	Yes	Yes

Note. See Exhibit 2 for definitions of symbols.

V. PRIME Applications and "Measurement Fads"

Because the pursuit of good measurement and evaluation is an active one in the course of Personnel Research, we should expect that the PRIME instrumentalities should enjoy wide utility.

In the measurement field there exists the tendency toward "fads." We have already seen a long period of the fad of "Norm-referenced measurement" wherein test items and tests are calibrated relative to a sample of persons from populations of interest.

We are presently witnessing the beginning of the period of the fad of "Criterion-Referenced measurement" wherein item and test calibrations are not so important, rather standards of a priori achievement are established and persons and groups of persons are compared to these standards.

"Individual-Centered measurement" is a fad which is merely starting and whose flood we can expect in the not too distant future. Concern here is for:

(a) The fit of item responses to the measurement model used.

(b) Items whose model parameters are invariant from population to population.

(c) Person ability or competency parameters of the model to exhibit invariance with respect to the items of the collection used to measure persons.

In spite of numerous attempts by others to achieve properties (b) and (c) above only the Rasch BMM model, used in the PRIME system, has these properties and the FRM program estimates the model parameters more simply and efficiently of the several methods available to do this. Thus the BMM-FRM system becomes even more important and unique as measurement procedures for Personnel Research work.

Exhibit 6 indicates which PRIME instrumentalities are likely to be useful tools of research in each of nine research program areas at NPRDC. Also specified are those variables which are required to be measured or evaluated and those variables for which values or utilities need to be assessed.

Let us examine one area in some detail, namely program 7, Measure Job Performance. New attempts at measuring on-the-job performance have never been very successful nor frequent. Indeed this is a difficult research area and useful tools and models are rare. We do know, however, that relevant data are sometimes at hand. For example, McDowell (1972) has collected "task" data from large samples of personnel in several Navy ratings. These data provide "level of quality" responses

for each of several hundred rating-related tasks and therefore could be used with PAM or PMM to obtain a measure of on-the-job competency for each questionnaire respondent. It appears that the "difficulty" in this area can, at least in part, be resolved by an effective interrelationship of the fields of Occupational Analysis and psychometrics. Each field seems to have already developed the appropriate technology to make the key contributions to solve the problem. The critical point being to use the categorical response data resulting from an occupational task survey together with an assessment model which can accommodate such data and subsequently derive a measure of person competency. These concepts have utility beyond research at NPRDC since Task Inventory data collection and analyses are being carried on elsewhere in the Navy (c.f. NOTAP) and by other services.

The assessment of "Effectiveness" of Persons, Systems or Objects seems of general importance. To use PRIME for this problem it appears that we must take the "bull by the horns" and indicate, for each application, the issues to be considered important for assessing effectiveness as well as the relative importance of each issue. The final say in these matters is of course the perogative of high level policy makers, but, at the research level, we can now provide a procedure and first attempt approximations to "effectiveness" policy. PRIME instrumentalities TIM and HELP seem to be appropriately relevant models, among others, for assessing effectiveness. We conclude this section by summarizing applications in Exhibit 6.

Summary

Our problem is called "Assessment" and our purposes were to develop a structure by which assessment models could be uniquely characterized and to develop assessment models and their associated computer programs which could be used for personnel research purposes.

To these ends two aspects of assessment, namely measurement and evaluation were identified. Their distinction lying mainly in the process used to respond to the items of an assessment instrument. There are, however, often elements, or their combinations, which can be used to clarify the distinction.

The parameters of the assessment models were identified. These were variously combined and operationalized into either measurement or evaluation models. The intention of this process being to make evaluation more measurement-like by using and developing evaluation models, for polychotomous response information, which are patterned after Rasch latent-trait measurement models. Also developed were some special models for use in

EXHIBIT 6

Potentially Relevant PRIME Instrumentalities to NPRDC Program Areas

No.	Program Title	Important Assessment Variables	Important Values or Utilities	Potentiall PRIME Instru for Program		
				FRM	PAM	PM
1.	Acquisition and Initial Service	Performance	Recruit Assignments	*	*	*
2.	Career and Occupational Design	Job Performance Occupational Success		*	*	*
3.	Attitude and Motivational Research	Attitudes, Motivation Operational Readiness			*	*
4.	Personnel Effectiveness	Personnel Effectiveness			*	*
5.	Training Technology	General Evaluations				
6.	Test Training Systems	Training Systems		*		
7.	Measure Job Performance	Job Performance		*	*	*
8.	Human Performance		Human Effectiveness		*	*
9.	People and Organizational Management	Organizational Effectiveness, Motivation, Processes			*	*
10.	Management Systems		Organizational Structure			*
11.	Human Information Processing		Decision Making Methods and Values			*

EXHIBIT 6

tially Relevant PRIME Instrumentalities to NPRDC Program Areas

	Important Assessment Variables	Important Values or Utilities	Potentially Relevant PRIME Instrumentalities for Program Assessments				
			FRM	PAM	PMM	TIM	HELP
d e	Performance	Recruit Assignments	*	*	*	*	*
u- n	Job Performance Occupational Success		*	*	*	*	*
oti- rch	Attitudes, Motivation Operational Readiness			*	*	*	*
c-	Personnel Effectiveness			*	*	*	*
	General Evaluations					*	*
	Training Systems		*			*	*
	Job Performance		*	*	*		
nce		Human Effectiveness		*	*	*	*
an- gement	Organizational Effectiveness, Motivation, Processes			*	*	*	*
tems		Organizational Structure			*	*	*
ion		Decision Making Methods and Values			*	*	*

decision making paradigms and for assessment situations for which an estimation sample does or does not exist.

Analysis of these models and their associated data is provided for by a system of computer programs called PRIME. In all, five programs were developed, illustrated and documented. Four are now operating on the GA 1830 computer system at NPRDC in San Diego, California.

Applications of these models and computer programs to personnel research areas of the NPRDC research program was discussed.

VII. REFERENCES

Allerup, P., & Sorber, G. (1975). "The Rasch Model for Questionnaires." Report of the Danish Institute for Educational Research, Copenhagen, Denmark.

Andersen, E. B. (1972). "A Computer Program for Solving a Set of Conditional Maximum Likelihood Equations Arising in the Rasch Model for Questionnaires," (unpublished). Educational Testing Service, RM-72-6, Princeton, New Jersey.

Andersen, E. B. (1973). "Conditional Inference for Multiple-Choice Questionnaires," British Journal of Mathematics, Statistics and Psychology, 26:31-44.

Anderson, S. B. Ball, S., Murphy, R. T., & Rosenthal, E. J. (1973). "Anatomy of Evaluation," Educational Testing Service, Princeton, N.J.

Birnbaum, A. (1968). Some latent trait models and their use in influencing an examinee's ability. Part V of F. M. Lord and M. R. Novick, Statistical Theories of Mental Test Scores. Reading, Massachusetts: Addison-Wesley Publishing Company.

Gulliksen, H. (1950). Theory of Mental Tests. New York: John Wiley.

Lord, F. M. (1952). "A Theory of Test Scores." Psychometric Monograph, No. 7.

Lord, F. M., & Novick, M. R. (1968). Statistical Theories of Mental Test Scores. Reading, Massachusetts: Addison-Wesley Publishing Company.

McDowell, H. C. (1972). "New Concepts to Improve Training and Shipboard Manpower Utilization in the Engineering and Hull Ratings: Statistical Analysis." Navy Personnel Research and Development Center, San Diego, California, NPTRL SRR 72-26, June 1972.

Moonan, W. J. (1974). "COMET: Contributions to Objective Measurement and Evaluation of Trainee Competency," Proceedings of the 1974 MTA, Oklahoma City, Oklahoma.

Moonan, W. J. (1974). "CHAROSEL: A Computer Program for Qualitative Criterion Prediction Problems," Proceedings of the 1973 MTA, San Antonio, Texas.

Moonan, W. J., & Covher, M. (1975a). "FRM," Tools for Analysis and Decision, FACT DOCUMENT No. 32, Naval Personnel Research and Development Center, San Diego, California.

Moonan, W. J., & Covher, M. (1975b). "PAM," Tools for Analysis and Decision, FACT DOCUMENT No. 34, Naval Personnel Research and Development Center, San Diego, California.

Moonan, W. J., Southwick, P. C., & Covher, M. H. (1975c). "TIM," Tools for Analysis and Decision, FACT DOCUMENT No. 33, Navy Personnel Research and Development Center, San Diego, California.

Moonan, W. J., & Covher, M. H. (1975d). "HELP," Tools for Analysis and Decision, FACT DOCUMENT No. 37, Navy Personnel Research and Development Center, San Diego, California.

Moonan, W. J., & Covher, M. H. (1975e). "MAXCO," Tools for Analysis and Decision, FACT DOCUMENT No. 35, Navy Personnel Research and Development Center, San Diego, California.

Page, E. B. (1975). "Evaluating 'Evaluation'," The Review of Education, 1(i), 133-140.

Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1970). "An Individual-Centered Approach to Item Analyses with Two Categories of Answers." Unpublished, 1969 AERA Conference on Rasch Models, Los Angeles, California.